IB/05/050415

证　　明

本证明之附件是向本局提交的下列专利申请副本

申　请　日：　2004.02.24

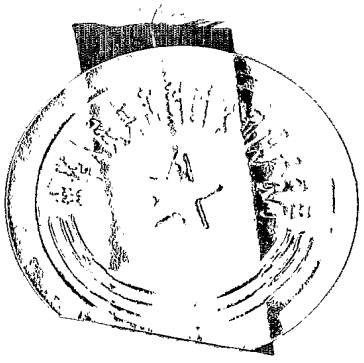
申　请　号：　200410007668.5

申请类别：　发明

发明创造名称：　一种节目内容定位方法和装置

申　请　人：　皇家飞利浦电子股份有限公司

发明人或设计人：　陈鑫、曾勇勤、陈宁江



中华人民共和国
国家知识产权局局长　王景川

2005 年 1 月 14 日

权 利 要 求 书

1．一种在一个多媒体节目中进行定位的方法，该多媒体节目包含一个带有文字信息的流，包括步骤：

a. 接收一个来自用户的请求, 该请求包含特定文字；

b. 确定所述的特定文字在所述的带有文字信息的流中的一个出现位置； 和

c. 确定与该出现位置处的文字信息同步的其他可呈现信息。

2．如权利要求 1 所述的方法，还包括步骤：呈现所述出现位置处的节目内容给用户。

3．如权利要求 1 所述的方法，其中所述的其他可呈现信息包括音频和视频中至少一种。

4．如权利要求 1 所述的方法，其中所述的文字信息是以文本形式存在的。

5．如权利要求 4 所述的方法，其中所述的其他可呈现信息包括图像。

6．如权利要求 1 所述的方法，所述的文字信息是以图像形式存在的，还包括步骤：

获取所述的文字信息对应的文本信息。

7．如权利要求 1 所述的方法，其中所述的带有文字信息的流中的内容具有层次性，还包括步骤：

确定一个包含所述出现位置的层，该层有一个特定的开始位置和一个特定的结束位置，从而使得步骤 c 中所确定的其他可呈现信息具有相应的开始位置和结束位置。

8．如权利要求 1 所述的方法，其中所述的来自用户的请求还包含一个开始位置信息和一个结束位置信息，该开始位置信息和结束位置信息是相对于所述的出现位置的，从而使得步骤 c 中所确定的其他可呈现信息具有相应的开始位置和结束位置。

9．如权利要求 7 或 8 所述的方法，还包括步骤：提取一个节目片段，该节目片段具有所述的开始位置和结束位置。

10．如权利要求 9 所述的方法，其中所述的多媒体节目是通过 SMIL 来集成的，所述的提取步骤是通过修改该多媒体节目的 SMIL 描述文件来完成的。

11．一种在一个多媒体节目中进行定位的装置，该多媒体节目包含一个带有文字信息的流，包括：

一个请求接收装置， 用于接收一个来自用户的请求, 该请求包含特定文字；

一个文本定位装置，用于确定所述的特定文字在所述的带有文字信息的流中的一个出现位置；和

12. 如权利要求 11 所述的装置，还包括一个呈现装置，用于呈现所述出现位置处的节目内容给用户。

13. 如权利要求 11 所述的装置，其中所述的其他可呈现信息包括音频和视频中至少一种。

14. 如权利要求 11 所述的装置，其中所述的文字信息是以文本形式存在的。

15. 如权利要求 14 所述的装置，其中所述的其他可呈现信息包括图像。

16. 如权利要求 11 所述的装置，其中所述的文字信息是以图像形式存在的，所述的文本定位装置还用于获取所述的文字信息对应的文本信息。

17. 如权利要求 11 所述的装置，其中所述的带有文字信息的流中的内容具有层次性，所述的文本定位装置还用于确定一个包含所述出现位置的层，该层有一个特定的开始位置和一个特定的结束位置，从而使得所述的同步定位装置所确定的其他可呈现信息具有相应的开始位置和结束位置。

18. 如权利要求 11 所述的装置，其中所述的来自用户的请求还包含一个开始位置信息和一个结束位置信息，该开始位置信息和结束位置信息是相对于所述的出现位置的，从而使得所述的同步定位装置所确定的其他可呈现信息具有相应的开始位置和结束位置。

19. 一种多媒体节目播放装置，包括：

一个内容接收装置，用于接收一个多媒体节目，该多媒体节目包含一个带有文字信息的流；

一个呈现装置，用于呈现接收到的多媒体节目给用户；和

一个定位装置，该定位装置包括：

一个请求接收装置， 用于接收一个来自用户的请求, 该请求包含特定文字；

一个文本定位装置，用于确定所述的特定文字在所述的带有文字信息的流中的一个出现位置； 和

一个同步定位装置，用于确定与该出现位置处的文本信息同步的其他可呈现信息。

20. 如权利要求 19 所述的装置，还包括一个提取装置，用于从所述的多媒体节目中提取一个特定片段。

说　明　书

一种节目内容定位方法和装置

背景技术

本发明涉及一种节目内容的定位方法和装置，尤其涉及一种在多媒体节目中根据内容来定位的方法和装置。

多媒体节目一般除了包括一个视频（Video）流和一个音频（Audio）流外，大多还包括一个图像（Image）流和/或一个文本（Text）流，这些流按照特定的规则和预定的时间顺序同步在一起，以供用户欣赏。在众多的多媒体节目的编排规则中，同步多媒体综合语言 SMIL（Synchronized Multimedia Integration Language）是一种流行的编辑语言。SMIL 不仅可以按时间顺序来整合多媒体节目的各个内容流，而且还可以用于管理该多媒体节目在呈现时的布局。

在多媒体节目的观看过程中，用户往往需要在该节目中找到一个特定片段，比如：在布什总统在清华大学的演讲这样一个多媒体演讲节目中，一位用户需要找到关于伊拉克问题的部分，对于该需求，用户可通过快进/快退的方式对音频的进行辨别，从而在节目中进行定位。再比如，在一部关于澳洲风光的多媒体记录节目中，一位用户希望直接浏览有关悉尼大剧院的片段，对于该需求，该多媒体播放装置可对视频流中的视频进行自动匹配分析，当画面中出现悉尼大剧院时，则呈现该片段给用户。

在上述的内容定位过程中，如果由用户来进行人工操作，则往往需要反复搜索多次才能找到所需的位置，浪费了用户的时间，给用户带来了很大的麻烦；如果由多媒体播放装置来进行自动搜索，则由于视频和音频的复杂性，视频流和音频流的数据量非常大，因此在其中进行搜索的难度也就很大，而且对硬件的要求也会比较高，从而增加了用户的成本。

另外，为了方便对多媒体节目，特别是多媒体演示节目的编辑，市场上出现了各种各样的制作工具（authoring tool），如美国 Accordent 公司的 PresenterOne 和加拿大 SofTV.net 公司的 Presentation Maker 等，这些工具可以让用户将一个多媒体演示的文本幻灯片的标题列在一个表里，用户可用这些标题作为索引来找到对应的片段。这样虽然一定程度上简化了一部分搜索过程，但条件是在该多媒体演示节目制作时需使用上述专业编辑工具。进一步说，

该编辑工具仅提供非常有限的标题供用户选择，这限制了用户选择的任意性，从而不能实现用户的个性化选择。

因此，需要一种新的节目内容定位方法和装置，可以使用户能够在一个多媒体节目中方便地进行内容定位，从而获得其所需要的特定片段，满足其个性化需求。

发明内容

本发明的目的之一是为了消除现有的节目内容定位方法和装置的缺陷，提供一种新的节目内容定位方法和装置，可以使用户能够在一个多媒体节目中方便地进行内容定位，从而获得其所需要的特定片段。

本发明提供了一种在一个多媒体节目中进行内容定位的方法，该多媒体节目包含一个带有文字信息的流。首先，接收一个来自用户的请求，该请求包含特定文字；其次，确定所述的特定文字在所述的带有文字信息的流中的一个出现位置；最后，确定与该出现位置处的文字信息同步的其他可呈现信息。所述的其他可呈现信息可以是视频信息，也可以是音频信息。

所述的文字信息可以是以文本形式存在的，也可以是以图像形式存在的。当其以图像形式存在时，该定位方法还包括步骤：获取该文字信息对应的文本信息。

所述的带有文字信息的流可以具有层次性，此时，该定位方法还包括步骤：确定一个包含所述出现位置的层，该层有一个特定的开始位置和一个特定的结束位置，从而使得最后所确定的其他可呈现信息具有相应的开始位置和结束位置。

本发明还提供了一种在一个多媒体节目中进行内容定位的装置，该多媒体节目包含一个带有文字信息的流，该文字信息可以是以文本形式存在的，也可以是以图像形式存在的。该装置包括一个请求接收装置、一个文本定位装置和一个同步定位装置。

该请求接收装置用于接收一个来自用户的请求，该请求包含特定文字；该文本定位装置用于确定所述特定文字在所述的带有文字信息的流中的一个出现位置；该同步定位装置用于确定与该出现位置处的文字信息同步的其他可呈现信息。所述的其他可呈现信息可以是视频信息，也可以是音频信息。

本发明通过对一个多媒体节目自身所包含的带有文字信息的流进行分析，定位出用户所需的节目片段位置，然后通过同步规则找到相应的视频或音频片段。由于相对于视频或音

频而言，带有文字信息的流，如文本流或图像流，所含数据量要少很多，同时对文本进行分析也大大简单于对画面或声音的分析，因此，本发明极大地简化了进行节目内容搜索的复杂性，降低了对硬件的要求，方便了用户的操作，满足了其个性化需求。

通过参照结合附图所进行的如下描述和权利要求，本发明的其它目的和成就将是显而易见的，并对本发明也会有更为全面的理解。

附图说明

本发明通过实例的方式，参照附图进行详尽的解释，其中：

图 1 是根据本发明的一个实施例的一个在一个多媒体节目中进行内容定位的装置的系统框图；

图 2 是根据本发明的一个实施例的一个在一个多媒体节目中进行内容定位的流程示意图；

图 3 是根据本发明的另一个实施例的一个在一个多媒体节目中进行内容定位并提取特定片段的流程示意图。

在所有的附图中，相同的参照数字表示相似的或相同的特征和功能。

具体实施方式

图 1 是根据本发明的一个实施例的一个在一个多媒体节目中进行内容定位的装置的系统框图。该装置 100 可以为一个多媒体节目制作装置（图中未显示）或一个多媒体播放装置（图中未显示）的一部分。装置 100 包括一个请求接收装置 120、一个文本定位装置 130 和一个同步定位装置 140。装置 100 还包括一个内容接收装置 110、一个呈现装置 150 和一个提取装置 160。装置 100 所包括的上述装置对于本领域的熟练技术人员来说可以通过多种现有的装置来实现，只要其组合在一起可以达到本发明的功能即可。

内容接收装置 110 用于接收一个多媒体节目，该多媒体节目包含一个带有文字信息的流，如文本流或包含有文字信息的图像流（在现有的多媒体演示节目中，作为演示辅助工具的幻灯片，比如，Powerpoint 文件中的一个页面，往往以图像形式来传输）。该多媒体节目可以来自于一个本地的存储装置（图中未显示），如 DVD；亦可来自于一个网络服务器（图中未显示）。

请求接收装置 120 用于接收一个来自用户的请求，该请求包含特定文字，如“悉尼大剧院”等，用户希望通过该请求在正在编辑/欣赏的多媒体节目中来找到介绍悉尼大剧院的片段，该多媒体节目包含一个带有文字信息的流。

文本定位装置 130 用于确定所述的特定文字在所述的多媒体节目中的一个出现位置。装置 130 在所述的带有文字信息的流中搜索该特定文字，如“悉尼大剧院”，在找到该特定文字后获得其在节目中的位置信息。如果前述的带有文字信息的流为一个图像流，装置 130 还用于获取该图像流中的文字信息对应的文本信息。

同步定位装置 140 用于确定与所述出现位置处的文字信息同步的其他可呈现信息。由于多媒体节目中不同内容流在时间上的同步性，因此可根据一个内容流，如文本流中的一个位置信息，确定该位置在其他内容流，如视频流或音频流中的相应位置。

呈现装置 150，用于呈现在一个多媒体节目中的一个特定位置的节目内容给用户。

提取装置 160，用于从一个多媒体节目中提取（extract）出一个特定片段，在本实施例中，该特定片段可包括前述的特定文本信息。

装置 100 的运行流程详见下面图 2 和图 3 所述。

图 2 是根据本发明的一个实施例的一个在一个多媒体节目中进行内容定位的流程示意图。首先，获取一个多媒体节目（步骤 S210），该多媒体节目包含一个带有文字信息的流，所述的文字信息以文本形式存在，比如，对于一个多媒体数字电视节目流，其中的字幕以文本形式存在于其数据流中；又如，对于一个多媒体演示节目，其演示的文字内容可以文本形式存在于一个文本流中。如果该多媒体节目较长，该步骤可以是一个持续的步骤，直到整个定位的流程结束为止。

本实施例中仍以关于澳洲风光的多媒体节目为例，该节目包含有一个文本流，在该流中包含有相应的解说词内容。

然后，接收用户的一个请求，该请求包含特定文字（步骤 S230），如“悉尼大剧院”，用户预期该特定文字会出现在上述的文本流的某一个位置，并希望通过该请求在步骤 S210 中所获取的多媒体节目中来找到包含有该特定文字的片段。

接下来，在上述文本流中搜索该特定文字，并判断是否找到该特定文字在所述的文本流中的一个特定出现位置（步骤 S230），如判断结果是否定的，则提示用户未能在该多媒体节目中找到所述的特定文字（步骤 S234），并结束整个流程；如判断结果为肯定的，则获取

该特定出现位置的信息（步骤 S238），如“悉尼大剧院”出现在距节目开始时“01：03：06”（hh:mm:ss）的位置。

再接下来，根据特定的多媒体节目的同步规则，确定该特定文字的出现位置在视频流中的相应位置（步骤 S240），如找到距节目开始时“01：03：06”（hh:mm:ss）位置的视频，该时刻处的画面通常包含有与解说词相对应的悉尼大剧院的景观。多媒体节目的同步规则可以有很多种，在此就不再一一列举。

最后，呈现该特定位置的视频给用户（步骤 S250），该处的画面包含有用户想要欣赏的悉尼大剧院的景观。当然，亦可呈现在该特定位置的多媒体节目的全部内容，如视/音频，图像和文本等，或其他部分内容，如音频，给用户，以满足用户的个性化需求。

在步骤 S250 的呈现过程中，还可以呈现该特定出现位置之前和/或之后的一段时间的视频。该时间的长度可通过用户设定时间值，或系统给定缺省值。用户可在步骤 S220 的请求中包含一个开始位置信息和一个结束位置信息，该开始位置信息和结束位置信息是相对于用户所预期的特定出现位置的。

当然，本实施例中，在步骤 S240 中，亦可根据同步规则，确定该特定文字的出现位置在音频流或图像流中的相应位置。因为无论视频或者音频，甚至图像，都要比文本复杂，对它们的分析及定位也都比对文本的分析及定位要复杂得多。由此可见，本发明所提出的定位方法比现有的通过音/视频来定位的方法要简单得多。

在上述定位过程中，如果特定文字，如“悉尼大剧院”在所述的文本流中多次出现，则可在步骤 S250 呈现特定位置的视频给用户的同时，给予用户一个选择是否继续搜索的机会，用户选择继续搜索，则从上一次搜索到的特定位置沿着原有的搜索方向继续搜索，直到找到用户想要欣赏的场景或节目结束。该选择机会可通过在屏幕上呈现一个按钮来提示用户是否需要继续搜索，然后接收用户的输入信息来完成。

图 3 是根据本发明的另一个实施例的一个在一个多媒体节目中进行内容定位并提取特定片段的流程示意图。首先，获取一个多媒体节目（步骤 S310），该多媒体节目包含一个带有文字信息的流，所述的文字信息以图像（image）形式存在，比如，对于一个多媒体演示节目，其演示的幻灯片（Slide）包含有文字信息内容，并以图像形式存在于一个图像流中。如果该多媒体节目较长，该步骤可以是一个持续的步骤，直到整个定位的流程结束为止。

表 1 为一个多媒体演示节目的 SMIL 描述文件（Script），该节目包含有一个视频流和一个与之同步的图像流，该图像流包含有该演示的幻灯片及其上的文字，这些文字以图像形式存在。

表 1： 一个多媒体演示节目

```
<smil>
 <head>
  <layout>
   <region id="r1" left="..." width="..." top="..." height="..." bottom="..." z-index="1"/>
   <region id="r2" left="..." width="..." top="..." height="..." bottom="..." z-index="1"/>
  </layout>
 </head>
 <body>
  <par>
   <video id="vid" region="r1" src="video_uri"/>
   <seq>
    <img id="img1" region="r2" dur="t1" src="image1_uri"/>
    <img id="img2" region="r2" dur="t2" src="image2_uri"/>
    <img id="img3" region="r2" dur="t3" src="image3_uri"/>
    <img id="img4" region="r2" dur="t4" src="image4_uri"/>
    <img id="img5" region="r2" dur="t5" src="image5_uri"/>
    <img id="img6" region="r2" dur="t6" src="image6_uri"/>
    <img id="img7" region="r2" dur="t7" src="image7_uri"/>
    <img id="img8" region="r2" dur="t8" src="image8_uri"/>
    <img id="img9" region="r2" dur="t9" src="image9_uri"/>
   </seq>
  </par>
 </body>
</smil>
```

从表 1 中可以看出，该图像流具有层次结构性，包含 9 个部分（image1、image2、image3、image4、image5、image6、image7、image8、image9），每一部分对应与一张幻灯片，亦即，每一部分有其特定的开始位置和延续长度。之所以如此，是因为在演示过程中一般视/音频在不断变化，而每一张幻灯片一般都要持续一段时间保持不变。

由于不能对以图像形式存在的文字直接进行文本分析，因此可通过一定方式获取该图像流中的文字信息对应的文本信息（步骤 S320），该获取步骤可运用现有的光学识别技术（OCR）来完成。

然后，接收用户的一个请求，该请求包含特定文字（步骤 S330），用户预期该特定文字会出现在上述的多媒体节目流的某一个或多个位置，并希望通过该请求来找到并提取包含有该特定文字的片段。

接下来，在上述图像流的文字信息中搜索该特定文字，并判断是否找到该特定文字的一个特定出现位置（步骤 S340），如判断结果是否定的，则提示用户未能在该多媒体节目中找到所述的特定文字（步骤 S344），并结束整个流程；如判断结果为肯定的，则获取该特定出现位置的信息（步骤 S350），比如：该特定文字出现 imge2 的文字信息中，则获取 image2 的开始位置及延续长度。

再接下来，根据特定的多媒体节目的同步规则，确定该特定文字的出现位置在视频流中的相应位置（步骤 S360），此时，相应视频流的特定片段的开始位置和延续长度与 image2 的开始位置和延续长度相同。

最后，根据获取的特定片段的开始位置和延续长度，修改原有的 SMIL 描述文件来得到一个新的 SMIL 描述文件（步骤 S370），该 SMIL 描述文件仅反映所找到的片段，从而实现了从该多媒体节目中提取出用户需要的特定片段。用户通过选择运行该修改后的 SMIL 描述文件可直接浏览其所需要的特定片段。

在步骤 S360 之后，还可进一步判断是否需要继续搜索（步骤 S380），如果判断结果是否定的，则结束整个提取流程；如果判断结果是肯定的，则回到步骤 S340 则从上一次搜索到的特定位置沿着原有的搜索方向继续搜索，直到找到用户想要欣赏的下一个片段或节目结束。该判断可通过判断该多媒体节目是否结束来自动进行，亦可通过给用户提示由用户来决定。

在本实施例中，除了上述在 image2 中找到所述的特定文本信息外，还在 image5 和 image8 中找到所述的特定文本信息，最终得到的修改后的 SMIL 描述文件如表 2 所示，该 SMIL 描述文件所对应的多媒体节目片段包含所述的特定文本信息。

表 2： 一个多媒体节目的特定片段

```
<smil>
 <head>
  <layout>
   <region id="r1" left="..." width="..." top="..." height="..." bottom="..." z-index="1"/>
   <region id="r2" left="..." width="..." top="..." height="..." bottom="..." z-index="1"/>
  </layout>
 </head>
 <body>
<par>
 <seq>
<video id="vid" region="r1" src="video_uri" clipBegin="smpte=T1" dur="t2"/>
<video id=" vid1" region="r1" src="video_uri" clipBegin="smpte=T2" dur="t5"/>
<video id="vid2" region="r1" src="video_uri" clipBegin="smpte=T3" dur="t8"/ >
</seq>
<seq>
 <!--non-highlight slides are removed-->
      <img id="img2" region="r2" dur="t2" src="image2_uri"/>
      <img id="img5" region="r2" dur="t5" src="image5_uri"/>
      <img id="img8" region="r2" dur="t8" src="image8_uri"/>
     </seq>
   </par>
  </body>
</smil>
```

其中：T1 = t1,

T2 = t1+t2+t3+t4

T3 = t1+t2+t3+t4+t5+t6+t7

本实施例中的多媒体节目的带有文字信息的流具有层次性，该层次性既可表现为上述的平行的仅有先后顺序的 9 个 image, 亦可表现为象书的章节一样，即不同的层次间可互相包含。

本发明由于利用了多媒体节目中本身所包含的带有文字信息的流进行定位，同时文字信息的分析相比于音/视频的分析要简单的多，因此，对于节目制作人员来说，可以节省大量的工作量，降低了工作的复杂性；对于用户来说，定位操作会变得相当方便，所需设备也相

对简单而且便宜。进一步说，还可通过语音识别（Voice Recognition）技术来将音频中的对白转换成文本来供定位之用。

虽然经过对本发明结合具体实施例进行描述，对于在本技术领域熟练的人士，根据上文的叙述作出的许多替代、修改和变化将是显而易见的。因此，当这样的替代、修改和变化落入附后的权利要求的精神和范围内时，应该被包括在本发明中。

说 明 书 附 图

100

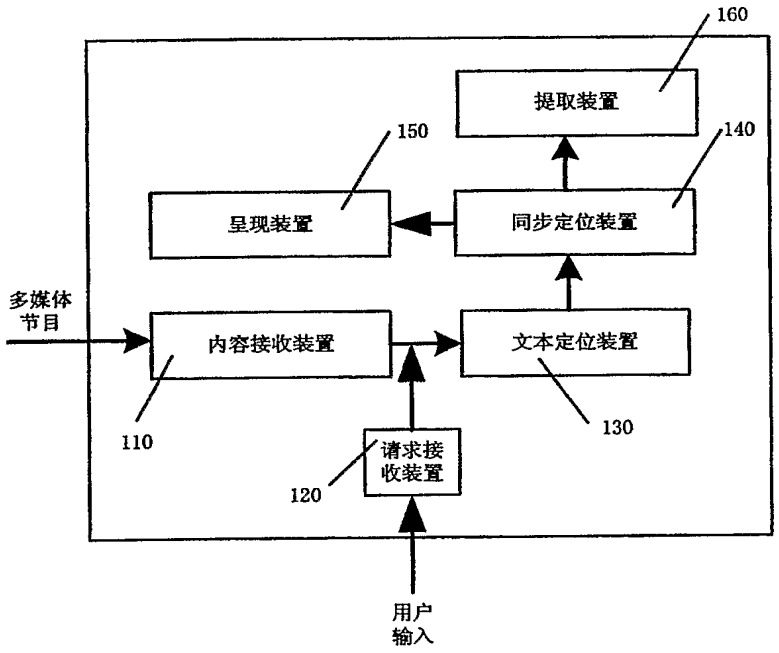

160
提取装置
150
140
呈现装置
同步定位装置
多媒体
节目
内容接收装置
文本定位装置
110
130
请求接
收装置
120
用户
输入

图 1

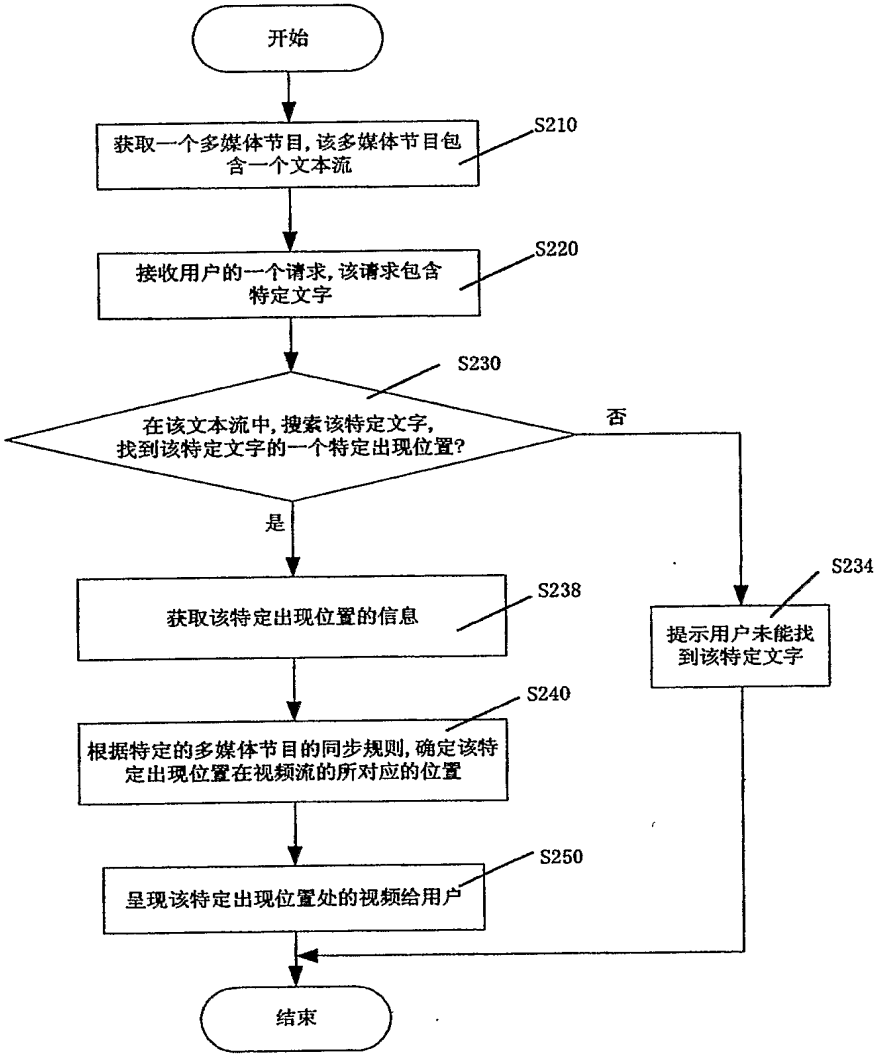

开始
获取一个多媒体节目，该多媒体节目包含一个文本流
S210
接收用户的一个请求，该请求包含特定文字
S220
在该文本流中，搜索该特定文字，找到该特定文字的一个特定出现位置？
S230
否
是
获取该特定出现位置的信息
S238
提示用户未能找到该特定文字
S234
根据特定的多媒体节目的同步规则，确定该特定出现位置在视频流的所对应的位置
S240
呈现该特定出现位置处的视频给用户
S250
结束

图 2

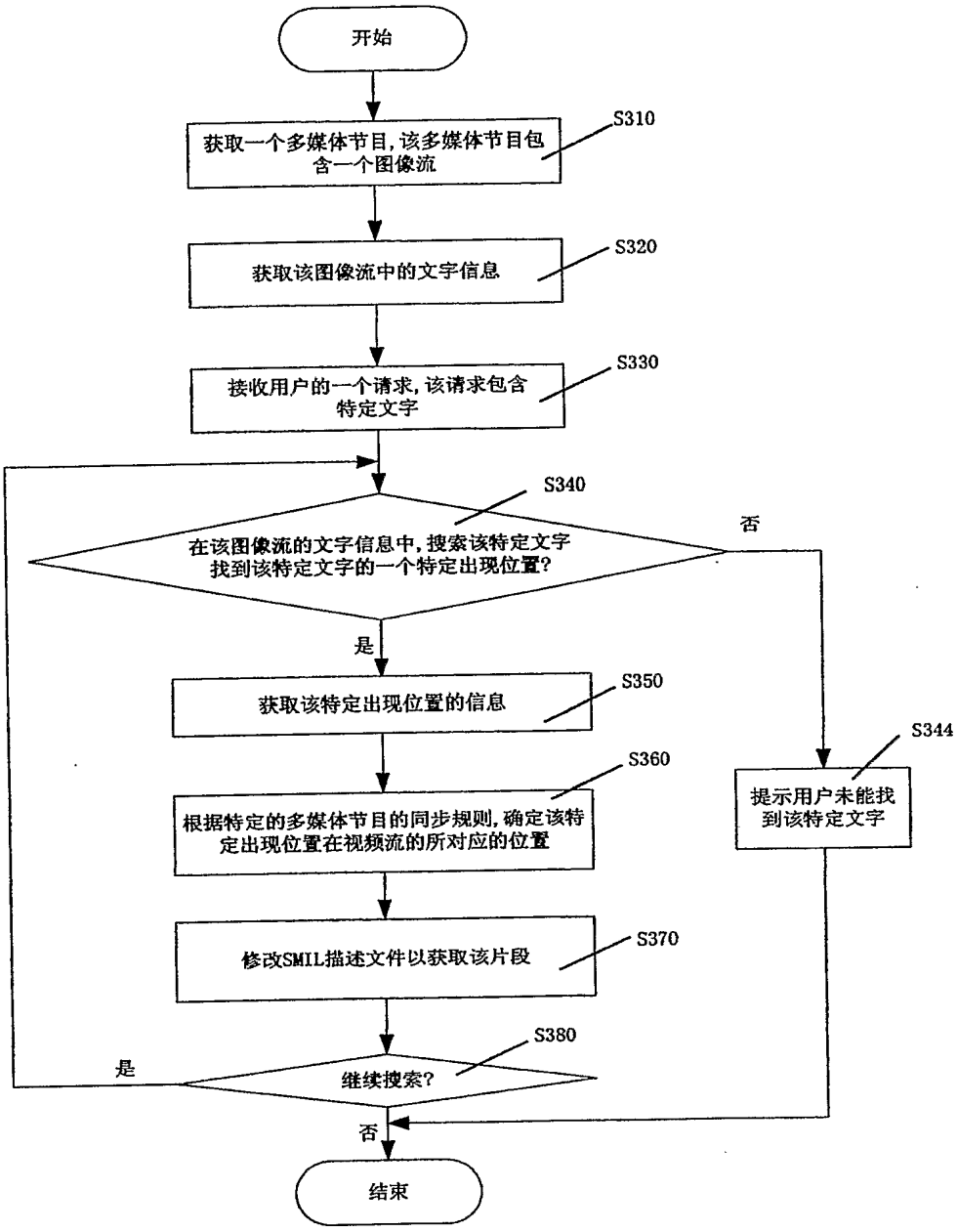

开始
获取一个多媒体节目，该多媒体节目包含一个图像流
S310
获取该图像流中的文字信息
S320
接收用户的一个请求，该请求包含特定文字
S330
在该图像流的文字信息中，搜索该特定文字找到该特定文字的一个特定出现位置？
S340
否
是
获取该特定出现位置的信息
S350
根据特定的多媒体节目的同步规则，确定该特定出现位置在视频流的所对应的位置
S360
提示用户未能找到该特定文字
S344
修改SMIL描述文件以获取该片段
S370
继续搜索？
S380
是
否
结束

图 3